Weather-induced crop failure events under climate change: a storyline approach

Henrique M.D. Goulart, Karin van der Wiel, Christian Folberth, Juraj Balkovic, and Bart van den Hurk

= Response to the reviewers =

Reviewer #3:

Review of "Weather-induced crop failure events under climate change: a storyline approach"

This study explored climate impacts on soybean crop failures in the Midwestern United States under present and future warming scenarios using the random forest model and storyline approach with model simulated crop data and CRU climate data. The findings suggest that the failures are likely to increase with global warming, changes in both univariate and multivariate climate drivers of failure decrease the importance of compound factor, and impact-analogues are significantly increased under global warming compared to event analogues. The manuscript is well in line with the scope of the journal Earth System Dynamics. While the random forest model is not new, it is novel to use that to analyze climate impacts on crop failures. Also, it is very interesting to use storyline approach to study this issue. The findings are relevant. I think, in general, the paper is publishable after revisions. While the statistical analysis is done rigorously, the data may be improved, and method description and the interpretation of the results and findings also need to be improved. Here I provide specific comments as follows.

We would like to thank the reviewer for their constructive feedback on our manuscript. We are grateful for the suggestions to improve our manuscript and we did our best efforts to address them in the comments below.

Major comments:

1. It is reasonable to use model simulated crop data to obtain longer time series for model training. However, the study does not show the performance of the simulated data compared to observations. Since the study region has accessible observational yield data over long period, it should demonstrate the simulated crop data can well represent observations. Otherwise, it may require a bias correction to improve the quality of the simulated crop data.

   Thank you for the suggestion. This has been pointed out also by another reviewer. We added a comparison between the EPIC-IIASA simulated yields and the observed yield dataset of the USDA (United States Department of Agriculture) for the region considered. First, we detrended linearly the observed dataset, then we spatially averaged both datasets along the period of 1960 to 2016 and finally we standardized the two datasets, to focus on the interannual variability of the yield timeseries. We found that the R^2 between EPIC and the observed dataset is high, at 0.674, and that EPIC is capable of replicating the interannual variability of the observed data.

   Line 115: "For validation of the crop model, we compared the EPIC-IIASA simulated yields with the observed yields from the US Department of Agriculture (USDA, www.nass.usda.gov/Quick_Stats) for the region

considered. EPIC-IIASA has higher mean and standard deviations values than the observed as the simulated yields are potential (Folberth et al., 2016).To evaluate the interannual variability, we obtained a coefficient of determination, R^2, of 0.674. We also observed a good correlation between the two standardised datasets (Figure C1). We consider EPIC capable of replicating the interannual variability of the observed data."

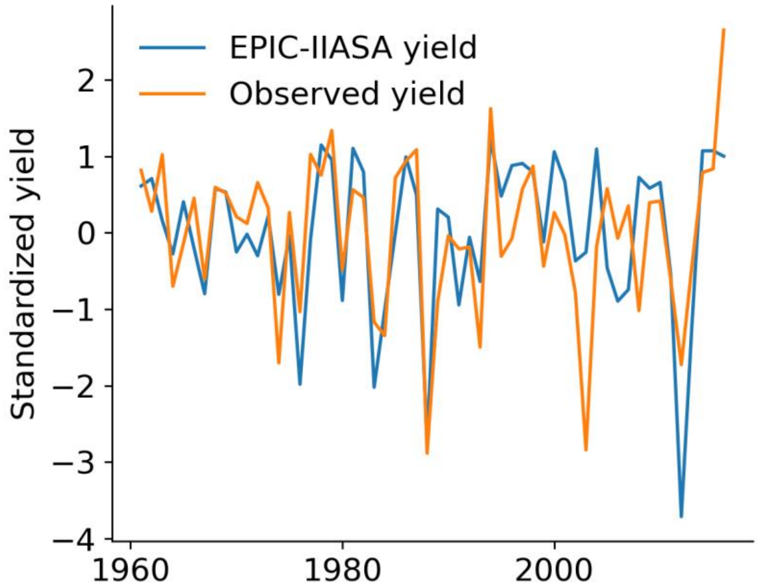


Figure C1: Standardized comparison between the EPIC-IIASA simulated yields and the observed yield dataset of the USDA (United StatesDepartment of Agriculture) for the region considered in this work.

2. It can be simple to average all the data over the region, but will discard many useful information. It is arbitrary to state "local scale of impacts are not meaningful for national and global implications". Using all the data points within the study region can increase the sample size by taking into account of the within region variability and increase the robustness of the model.

Thank you for the comment. We agree with the reviewer that aggregating data spatially might lead to loss of information, especially on local extreme conditions. However, the aggregation allows us to prioritise the dominant meteorological dynamics in the region, improving the clarity of the main interactions between climate and crop. As described in Section 2.2, there were preliminary steps to clean the data and minimise the spatial variability to obtain a more clear signal when aggregating across the region studied. We selected only the main soybean producing states and only grids with rainfed soybeans, obtaining a sub-region of the Midwestern US. In addition, there are other works on climate-crop interactions that have used different techniques of spatial aggregation from sub-national areas to global areas, like Lobell and Field, 2007; Lesk et al., 2016; Heino et al., 2018 and Ben-Ari et al., 2018. Since we agree with the reviewer that there are potential drawbacks in aggregating gridded data, we added to the discussion section a clarification on the limitation of the approach.:

Line 140: "We spatially averaged all data for the region studied (Figure 2) to focus on the regional scale of weather events and their crop yield impacts, as these have larger influence on global markets. Aggregating data spatially might lead to loss of information, especially on local extreme conditions, but the RF model performance is comparable when running on aggregated data and on all grid points (Table C1).'

Line 388: "We spatially aggregate the climate and crop data over the region analysed to focus on crop failures and meteorological conditions at the regional scale. While this approach allows us to focus on the main dynamics of the region, information on local extreme conditions is not attained."

3. It is not clear how the random forest model was trained and validated. While a description of the evaluation metrics is provided, it is not clear how the data was split as training data and validation data. The validation data needs to be excluded as unseen data when is used for model training.

We thank the reviewer for the suggestion. We added a sentence describing how the data was divided and how the model was trained and validated

Line 178: "With the RF setup complete, we trained and validated the RF model following the 80/20 split, where 80% of the data was used to train the model and the remaining 20% was used to validate the model's performance on unseen data."

4. All the results require thorough interpretations such as the mechanisms and biophysical meaning besides just reporting the data.

Thank you for the comment. We believe the scope of the work is more focused on the climatological dynamics of global warming, crop failures and the societal implications that these might bring. Therefore, to explain in detail the biophysical mechanisms of soybean development is not the focus. We have added extra discussion points on DTR (which is answered in comment 5), and on the interpretation of the compound events associated to soybean failures:

Line 428: "From a multivariate perspective, the correlation structure of the variables contributes to the occurrence of compound events, as previously shown by (van den Hurk et al., 2015) and (Santos et al., 2021). Yet, we observe a decrease in its importance under global warming conditions for the failure yield threshold adopted here (Figure 6). A higher frequency of years with critical temperature during summer makes crop failures mostly dependent on precipitation values. Therefore, while still physically a compound event, the soybean failures under global warming become statistically similar to a univariate event based mostly on precipitation."

5. Based on the previous literature, the increase in DTR may actually benefit yields, which is contradictory with the finding from this study. Please explain.

Thank you for the comment. Following the suggestion of the reviewer, we expanded on the literature about the influence of DTR on crop development. We found that DTR could be linked to both positive and negative impacts on crop yields, as high values of DTR could be a sign of increased heat or cold stress, and low DTR could be a consequence of high cloud coverage, low solar

radiation, or high rainfall. Therefore, we believe it is plausible that our statistical model found an association of crop failures with high values of DTR. We updated the text as:

Line 418: "The DTR projections indicate a descending trend, which in itself reduces crop failure probability according to our model. DTR is highly relevant for crop development, with previous studies showing the multiple impacts it can have on crops (Lobell, 2007; Zhang et al., 2013; Chen et al., 2015; Verón et al., 2015; Hernandez-Barrera et al., 2017; Rahman et al., 2017; van Etten et al., 2019). High values of DTR suggest peaks in high day-time temperature, which can disrupt the photosynthetic activity of crops (Allakhverdiev et al., 2008). High values of DTR can also indicate low night-time temperatures or night frosts, with capacity to damage crops during all stages of crop growth (Barlow et al., 2015). A study based on EPIC to simulate maize yields in the US has demonstrated that higher DTR values lead to greater evapotranspiration losses, reducing the yield outputs (Dhakhwa and Campbell, 1998). On the other hand, low values of DTR could indicate low solar radiation or high cloud coverage, both harmful for crops (van Etten et al., 2019; Vogel et al., 2019; Lobell, 2007). The decrease in DTR values due to global warming is mainly associated with a higher increase of minimum night-time temperature than daytime temperature (Qu et al., 2014; Sun et al., 2019). "

Minor comments:

1. The approach formulated in this study consists of three parts. The storyline approach is part 3 of the approach (part c) and address part of the research objectives. However, based on the title, storyline approach seems the key message of this study. I think the title needs to be changed since the storyline approach seems not the major focus of this study.

   While it is true that the last third of the work is the one dedicated to storylines, the chain of events that composes the storyline is developed in the first part (part a). We also believe the storylines are essential to the main message of this study: by creating a link between the meteorological conditions and the crop failure events, we were able to identify different meteorological conditions in the future projections that could lead to similar extreme-impact events to our case study. These different analogues following the climate-crop connection belong to the category of event-based storylines. Changing the title to "Storylines of weather-induced crop failure events under climate change" removes the focus on the storyline approach, but it expresses how we used storylines to create different analogues of a historical case.

2. Little result is presented in the abstract. It can be more informative to shorten the background and methods descriptions and include more results.

   We appreciate the suggestion. We added:

   Line 13: "We find that crop failures in the Midwestern US are linked to low precipitation levels, high temperature and diurnal temperature range (DTR) levels during July and August. Results suggest soybean failures are likely to increase with climate change. With more frequent warm years due to global warming, the joint hot-dry conditions leading to crop failures become mostly dependent on precipitation levels, reducing the importance of the relative

compound contribution. While event-analogues of the 2012 season are rare and not expected to increase, these analogues show a significant increase in occurrence frequency under global warming, but for different combinations of the meteorological drivers than experienced in 2012. This has implications for assessment of the drivers of extreme impact events."

3. Line 55: it is not clear to say "crop failures are compound events". Do you mean crop failures can be consequences of compound events? It requires a clear definition on compound events.

> ***Thank you for the correction. We mean crop failures are usually consequence of compound events. In line 36 we provide a definition of compound events. We updated the text as:***
>
> ***Line 59: "since crop failures are usually the result of compound meteorological drivers…"***

4. Lines 74-75: "…, a natural next step is to …" It may be not so natural. I suggest the authors better explain their motivations and why such study is needed.

> ***Thank you for the comment. We updated the section with a clearer sentence on the motivation for this work (1st quote) and added a more explicit sentence on the research question (2nd quote).***
>
> ***Line 78: "Building on these works, we expand the studies of global warming impacts on agriculture to include multivariate analysis, by explicitly modelling the compound nature of meteorological variables and their interactions."***
>
> ***Line 80: "The aim of this work is to understand how global warming affects the meteorological conditions leading to crop failures."***

5. The last paragraph of the introduction mostly described the method/experimental outline. The majority of this part should be better move to the method section as a general statement of the method prior to detailed descriptions. The introduction section should focus on explaining why the research is needed, is novel, and stating the scope, research questions/objectives.

> ***Thank you for the comment. Here we follow the convention to conclude the introduction section with a brief reading guide, the outline.***

6. Line 119: "… a similar 5-year period …" Please clarify what are the similar 5-year period at 2C and 3C.

> ***We updated the text to make the description clearer:***
>
> ***Line 131: "… a 5-year period representing an average global mean temperature 2°C above the pre-industrial levels (referred to as 2C) and another 5-year period corresponding to an average of 3°C above pre-industrial levels (referred to as 3C)"***

7. Line 194: Please interpret what is compound factor besides providing the equation.

> ***Thank you for the comment. Another reviewer also had questions on the compound factor, and we decided to reformulate the compound factor to relative compound contribution, in order to avoid confusion with compound***

events. The relative compound contribution quantifies how important the correlation structure of the meteorological variables is to crop failures, and how this might change with different levels of global warming. The interpretation of this metric is that at lower values of relative compound contribution, the order of the distribution of the meteorological variables becomes less important, which reduces the uncertainty of the future impacts with respect to the distribution of the meteorological variables. From a statistical point of view, it means the failure function becomes more dependent on a single variable, approximating to a univariate distribution. We updated the text as:

Line 209: "To assess the importance of the correlation of the conditions leading to failures, we created permuted versions of each large ensemble by randomly reshuffling the meteorological variables (van den Hurk et al., 2015; Santos et al.,2021), so that the correlation structure between them was removed (referred to as shuffled versions). We also defined a metric called the relative compound contribution. Relative compound contribution measures the importance of the correlation structure between the meteorological variables leading to crop failures. It is a statistical interpretation comparing crop failures under different correlation structures. Relative compound contribution is calculated as the ratio of the failure ratio obtained with the original data to the failure ratio obtained with the shuffled data. The closer to 1 the relative compound contribution gets, the less important the correlation structure between the variables is."

8. Lines 196-197: the definition of return period is confusing. Based on Figures 5 and 6, return period is not a one to one relationship with inverse of the failure probability, how it can be calculated as the inverse of the failure probability? It needs to be clarified.

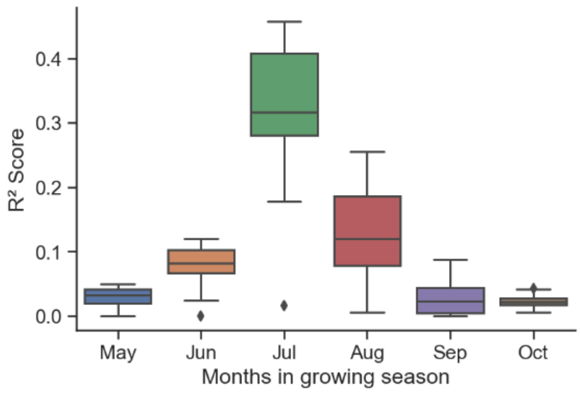
Thank you for showing us the ambiguity of the text. We corrected it. The return period is the inverse of the exceedance probability of an event to occur. In our study, we ordered the crop failure probabilities in an ascending order, calculated the exceedance probability and only then calculated the inverse of it to generate the return periods. So, the return period is not the inverse of the crop failure probability as the reviewer has correctly pointed out.

Line 219: "… comparing the failure probabilities for different return periods (calculated as the inverse of the exceedance probability of the specific event to occur) …"

9. It is not clear which month of data was used as candidate meteorological input prior to eliminating and combining monthly data in July and August.

We added a figure in the SI to illustrate the R^2 scores of the meteorological variables grouped by month, and added a short description:

Line 248: "The considered soybean season in the US ranges from May to October, and we see that the months with the highest sensitivity to meteorological conditions are July and August (Figure C6)."



10. Variable names and acronyms are inconsistent in the text and tables and do not follow standard practice.

Thank you for the suggestions. We updated the variable names and acronyms.

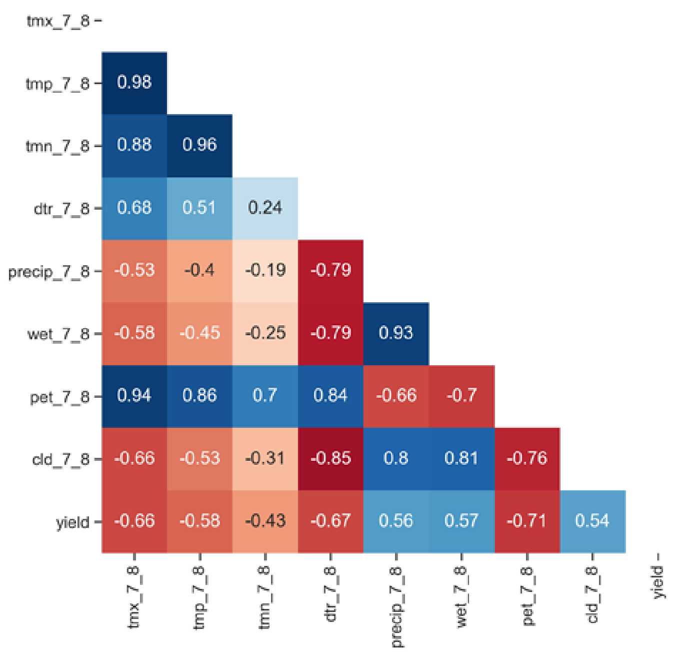
11. Figure 3b: Is Pearson's correlation performed on detrended or categorical yield?

Thank you for the question. The Pearson's correlation was performed on continuous detrended yield datasets.

12. I would suggest only providing half of the correlation matrix. To limit redundancy

Thank you for the suggestion, we updated the matrix.

b)



13. Figure 4: I am assuming that the ticks on the bottom of the partial dependency plot represent the percentiles of the data. If they are being included, they need to be explained.

Thank you for the comment. The other reviewers also questioned the purpose of the decile marks (ticks) and we decided to remove them, as they are not very informative. In addition, we added the full names of the variables as suggested.

14. Ln 390 decreasing compound factor. Is this just an artifact of choosing DTR.

This is related to a comment by the other reviewer. The decreasing relative compound contribution is related to the increasing temperature levels under global warming scenarios. With more years warm, the crop failures become statistically more similar to a univariate distribution.

Line 428: "From a multivariate perspective, even though the correlation structure of the variables still contributes to compound events, as previously shown by (van den Hurket al., 2015) and (Santos et al., 2021), we observe a decrease in its importance under global warming conditions (Figure 6). A higher frequency of years with critical temperature during summer makes crop failures mostly dependent on precipitation values. Therefore, while still physically a compound event, the soybean failures under global warming become statistically similar to a univariate event based on precipitation"

15. Ln 374 Figure 10d?

Thank you for the comment, it has been updated.